УДК 004.934

О.А. Савенкова, О.Н. Карпов
Днепропетровский национальный университет, г. Днепропетровск, Украина

Технология построения интеллектуальной системы распознавания речи

Предложена информационная технология для системы распознавания речи с большим словарем на основе алгоритма сегментно-слогового синтеза траектории параметров. В основу технологии положены модели обучения и распознавания, использующие быстросходящиеся алгоритмы поиска.

Введение

Создание компьютерных интерфейсов с речевым вводом/выводом информации требует решения проблемы построения системы распознавания речи (СРР) с большим словарем, для чего необходимо: определить структуры объектов распознавания и операции, необходимые для интеллектуального решения проблемы; разработать быстросходящиеся стратегии для эффективного поиска потенциальных решений, которые могут быть сгенерированы этими структурами и операциями с учетом дополнительной информации (эвристик) об исследуемой проблемной области [1-10].

Существующие технологии построения СРР с большим словарем используют подход, при котором предполагается, что в предъявленном речевом сигнале (РС) последовательность сегментов рассматривается как совокупность независимых событий, которым в соответствие ставятся независимые речевые единицы (РЕ) [3-6]. Однако такой подход не учитывает зависимости между параметрами смежных сегментов. Необходимо разработать информационную технологию, которая оперирует с непрерывными траекториями параметров и позволяет аппроксимировать любую реализацию РС в терминах РЕ как функцию времени. В качестве РЕ, которые могут использоваться в процессе синтеза (композиции) слов или предложений, удовлетворяют требованию максимальной полноты покрытия множества слов и учитывают зависимости между сегментами, могут быть выбраны слоги [7], [11], [12].

Постановка задачи

Разработать информационную технологию построения СРР с большим словарем на основе алгоритмов реализации сегментно-слогового синтеза траекторий параметров и ее программную реализацию.

Основная часть

Разработка информационной технологии требует решения следующих важных задач, которые вытекают из перечисленных актуальных проблем.

1. Создание оптимальных словарей РЕ: выбор способа хранения информативных параметров РЕ словаря.
2. Реализация автоматической сегментации РС.

3. Реализация эффективных алгоритмов распознавания на основе сегментно-слогового синтеза последовательности параметров для предъявленного РС.

В соответствии с иерархическими принципами построения СРР [11], результатами экспериментальных исследований [13], [14], [15] и с учетом рекомендаций к проектированию и разработке информационной технологии [16], в основу решения поставленной задачи положим две модели: модель обучения и модель распознавания.

1. Модель обучения

Общая схема модели обучения данной информационной технологии представлена на рис. 1 и состоит из следующих процедур. Рассмотрим более детально основные процедуры модели обучения.

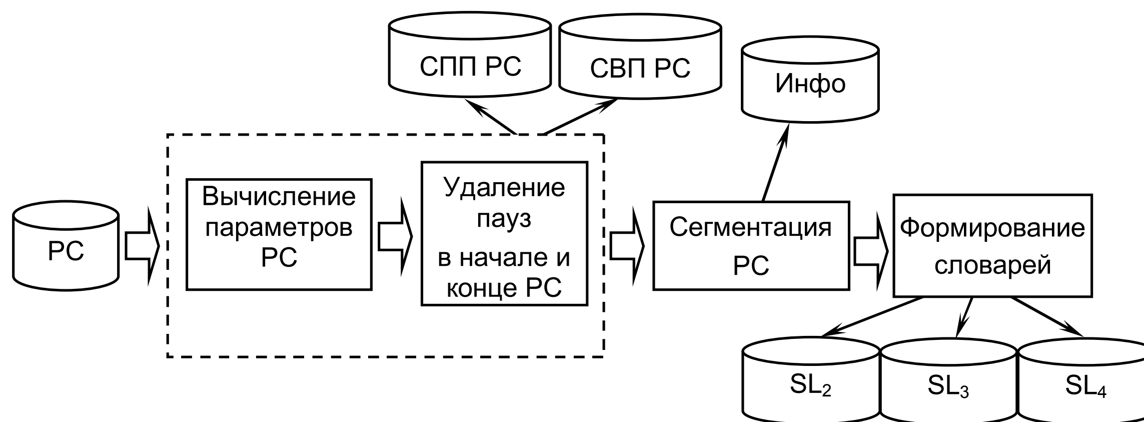


Рисунок 1 – Функционально-структурная схема модели обучения СРР

1. Ввод речевого сигнала с микрофона или открытие wav-файла с необходимой записью речевого сообщения.

2. Первичная обработка РС.

Первичная обработка РС заключается в вычислении информативных параметров РС (формировании траектории параметров) и удалении пауз в начале и конце речевого высказывания.

Процедура формирования ТП реализации РС: на каждом i-м интервале анализа РС вычисляется *вектор признаков* (набор информативных параметров) $x_i = \{x_{i1}, x_{i2}, \ldots, x_{ij}, \ldots, x_{in}\}$, далее формируется *траектория параметров* (ТП) анализируемого РС, таким образом $X = \{x_1, x_2, \ldots, x_i, \ldots, x_p\}$,

где p – число интервалов анализа РС; n – размерность вектора признаков.

Для представления ТП X исследуемого РС в данной информационной технологии выбраны [11], [13]:

– спектрально-временное представление (СВП) $XA(\omega, t)$;

– спектрально-полосное представление (СПП) $XE(l, t)$ в 9 частотных полосах.

СВП и СПП выбраны из следующих соображений: проведенные в [13] исследования по распознаванию РС, параметры которых представлены СВП и СПП, по критериям быстродействия и надежности показали, что увеличение быстродействия распознавания в случае использования СПП в среднем в 9 раз приводит к увеличению ошибки распознавания на 1,6 % по сравнению с использованием СВП.

Таким образом, по надежности распознавания оба метода представления ТП практически сравнимы и могут совместно использоваться как иерархия распознавания: СПП для предварительного распознавания и выделения списка кандидатов на распознавание, а затем СВП для принятия окончательного решения.

Процедура удаления пауз в начале и конце речевого высказывания. В качестве критериев нахождения границ сегментов, содержащих паузы в начале и конце высказывания, используются значения динамически меняющихся порогов нормированных энергий в двух полосах: 86 – 3010 Гц (НЭ); 3956 – 8084 Гц (ВЭ). Для пауз характерны минимальные значения энергий на всем частотном диапазоне: $ВЭ \le ВЭ_{min}$, $НЭ \le НЭ_{min}$, где $ВЭ_{min}$, $НЭ_{min} \in [0{,}03,\ 0{,}06]$.

3. Автоматическая сегментация РС.

Сегментация РС выполняется независимо для СВП и СПП по методу верификации временной последовательности параметров, при этом уточняются границы сегментов-фонем [11].

4. Формирование словарей для речевых единиц.

Для формирования словаря РЕ выполняется автоматическое деление сегментированной речевой последовательности параметров на акустические слоги-эталоны (двух-, трех-, четырехсегментные слоги) с соответствующей маркировкой лингвистической информации (название РЕ, транскрипция). ТП слогов-эталонов и вспомогательная информация распределяются в словари в соответствии с форматом хранения данных. Алгоритм автоматической декомпозиции РС на слоги-эталоны приведен на рис. 2.

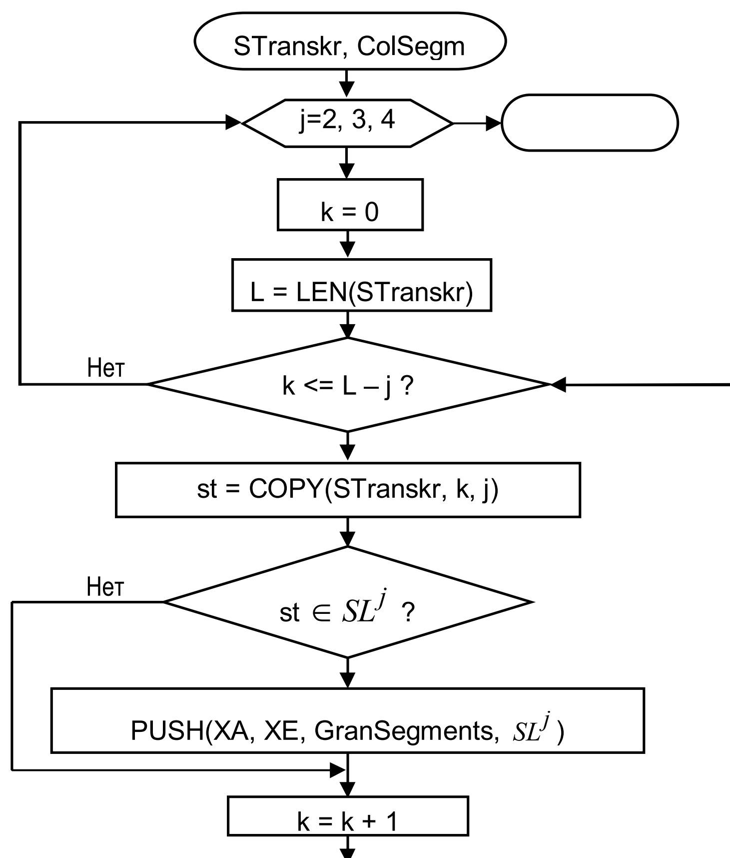


Рисунок 2 – Алгоритм автоматической декомпозиции РС на слоги-эталоны

Информация о речевой единице $\mathrm{SL_k}$ ($k = 1 \div N_{SL}$) в словаре представлена в следующем виде:

<*Номер РЕ*>_<*Имя РЕ*>_<*Транскрипция РЕ*>_
<*Количество временных отсчетов*>_<*Количество сегментов*>_
<*Адреса границ сегментов*>_
<*Групповая принадлежность сегментов*>.

Для создания СРР, которая работает в реальном времени, возникает необходимость создания оптимальных словарей РЕ с точки зрения объема занимаемой памяти и быстродействия их обработки. Следовательно, необходимо получить для каждого способа представления ТП модель описания с помощью аналитических функций, согласно которой можно восстановить исходную ТП с минимальной погрешностью.

Сплайн-описание СПП речевых единиц словаря. После сглаживания ТП в полосах на каждом временном участке, который соответствует сегменту-фонеме, последовательности параметров СПП $YE(l,t)$ в каждой частотной полосе имеют простую форму, которую можно описать полиномами низких порядков ($n \le 3$) [11]. Таким образом, для каждой k-й РЕ словаря может быть построена следующая модель сплайн-описания сегментированной ТП $YE(l,t)$, которая с достаточной точностью аппроксимирует последовательности параметров в каждой частотной полосе l ($l = 1 \div 9$)

$$\tilde{\mathrm{YE}} = \begin{cases} \tilde{YE}_{1\,i}, & s_0 \le i \le s_1, \\ \tilde{YE}_{2\,i}, & s_1 \le i \le s_2, \\ \ldots \\ \tilde{YE}_{k\,i}, & s_{k-1} \le i \le s_k, \\ \ldots \\ \tilde{YE}_{N_{SG}\,i}, & s_{N_{SG}-1} \le i \le s_{N_{SG}}, \end{cases} \qquad (1)$$

для каждого сегмента находятся параметры полинома 3-го порядка

$$\tilde{YE}^i_{k,l} = a^i_{k,l} \cdot (t - s_i)^3 + b^i_{k,l} \cdot (t - s_i)^2 + c^i_{k,l} \cdot (t - s_i) + d^i_{k,l}\,, \qquad (2)$$

где $k = 1 \div N_{SG}$ (N_{SG} – количество сегментов в ТП слога-эталона); $s_0, s_1, \ldots, s_{N_{SG}}$ – границы сегментов внутри слога-эталона $YE(l,t)$.

Для нахождения неизвестных параметров сплайн-описания решается задача минимизации среднеквадратического приближения с условиями в точках сегментации, которые обеспечивают требуемую гладкость склеивания ТП сегментов в каждой частотной полосе l (точки сегментации являются узлами сплайна)

$$y_l{}^2 = \sum_{j=1}^{N_{SG}} \left(\sum_{i=s_{j-1}}^{s_j} \left| \tilde{YE}^k_{i,l} - YE^k_{i,l} \right|^2 \right) \to \min\,, \qquad (3)$$

где s_j ($j = 1 \div N_{SG}$) – границы сегментации, N_{SG} – количество сегментов для k-й РЕ словаря; Y^k - исходные траектории параметров слогов-эталонов $YE(l,t)$, $\tilde{Y}^k$ – траектории параметров слогов-эталонов, восстановленные согласно модели (1), (2).

Описание СВП речевых единиц словаря в классе колоколообразных функций. Для построения аналитического описания СВП $YS^*(\omega, t)$ элементов словаря используется колоколообразная функция модифицированный локон Аньези, которая обладает следующими свойствами [17]:

1) функция гладкая непрерывная и дифференцируемая во всей области определения, имеет максимум на заданной координате и асимптотически приближается к плоскости области определения в любом направлении от максимума;

2) достаточно универсальная функция, форму которой можно варьировать в широких пределах с помощью изменения ее параметров a, b, c и имеет вид

$$Z(x) = \frac{a^3}{c^2 + (x-b)^2}. \tag{4}$$

По алгоритму, предложенному в работе [17], для элементов словаря SL_k, ТП $YA(\omega_k, t_l)$ которых представлены в некоторой частотно-временной области Ω: $[\omega_0,\ \omega_M] \times [t_0,\ t_N]$ (ω_k, t_l – дискретно заданные частота и время, $k = 1...M$, $l = 1...N$), вычисляются параметры колоколообразных функций $\{Zt(t_l)\}$, $\{Z\omega(\omega_k)\}$, которые обладают вышеперечисленными свойствами 1), 2), причем:

– функция $Zt_{(i)}(t_l)$ ($i = 1...L$) описывает временные свойства компонент речевого сигнала, определена в диапазоне $Zt \in [t_0, t_N]$;

– функция $Z\omega_{(i)}(\omega_k)$ ($i = 1...L$) описывает частотные свойства компонент речевого сигнала, определена в диапазоне $Z\omega \in [\omega_0, \omega_M]$.

Аналитическое описание $YA^*(\omega_k, t_l)$ СВП элемента словаря в некоторой точке области определения Ω вычисляется как суперпозиция L произведений колоколообразных функций $Zt_{(i)}(t_l)$, $Z\omega_{(i)}(\omega_k)$ ($k = 1...M$, $l = 1...N$, $i = 1...L$).

$$YA^*(\omega_k,\ t_l) = \sum_{i=1}^{L} Z\omega_{(i)}(\omega_k) \cdot Zt_{(i)}(t_l) = \sum_{i=1}^{L} \frac{a_{(i)}^3}{c_{(i)}^2 + (t_l - T_{(i)})^2} \cdot \frac{b_{(i)}^3}{d_{(i)}^2 + (\omega_k - \Omega_{(i)})^2}. \tag{5}$$

Таким образом, с учетом моделей аналитического описания ТП речевых единиц структура оптимального словаря может быть представлена на рис. 3.

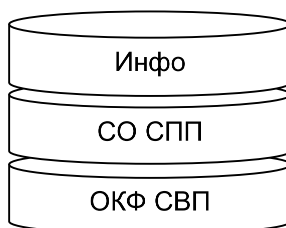


Рисунок 3 – Структура словаря речевых единиц SL_k

Инфо – лингвистическая информация о РЕ SL_k (название РЕ, транскрипция) + вспомогательная информация о РЕ SL_k (количество временных отсчетов, количество сегментов, границы сегментов, групповая принадлежность сегментов Тон-Шум-Пауза).

Двух- и трехсимвольные РЕ (слоги) SL_{2k}, $SL_{3k} \subset SL_k$, которым соответствуют двух-, трех- и четырехсегментные ТП ($\{YSL^{2k}\}$,$\{YSL^{3k}\}$,$\{YSL^{4k}\}$) представлены в виде:

СО СПП – параметров моделей сплайн-описания СПП;

ОКФ СВП – параметров моделей описания СВП в классе колоколообразных функций.

Анализ элементов словаря и вычисление эвристических оценок.

Наиболее значимые характеристики сегментно-слогового представления речевой последовательности:

– вложенность РЕ (слогов) $SL_{2k} \in SL_{3k}$, $SL_{2k} \in SL_{4k}$. Значение оценки вложенности h_v рассчитывается с учетом количества совпадений/несовпадений элементов лингвистической информации на соответствующих позициях и их порядка следования, $0 \le h_v(n) \le 1$;

– наличие определенной структуры групповых признаков сегментов (Т – «Тон»; Ш – «Шум»; П – «Пауза»). Значение оценки $h_g(n)$ рассчитывается на основе анализа количества сочетаний групповых признаков сегментов Т-Ш-П, $0 \le h_g(n) \le 1$;

– величина расстояния между ТП слогов, содержащих различные сочетания групповых признаков. Оценка расстояния $h_d(n)$ рассчитывается для возможных сочетаний сегментов, относящихся к разным групповым признакам, например, Т-Т, Ш-Т, Т-П.

На основе рассчитанных эвристических оценок формируется структура группировки элементов в словарях, благодаря чему сокращается время на поиск подходящего элемента при распознавании.

2. Модель распознавания

Сформулируем задачу построения модели распознавания на основе сегментно-слогового синтеза ТП согласно [11], [14], [15] таким образом.

Пусть задан словарь $\{SL_k\}$, состоящий из N слогов. Для каждого слога заданы эталонные последовательности параметров $\{Y_k\}$. Каждый слог SL_k сегментирован на n_k сегментов-фонем SG^y_{kj} ($k = 1 \div N$, $j = 1 \div n_k$).

Пусть задана входная последовательность параметров X, которая сегментирована на m_p сегментов-фонем SG^x_{pi}, объединяемых в M групп-слогов (сочетания двух, трех или четырех сегментов) SL^x_p ($p = 1 \div M$, $i = 1 \div m_p$).

Необходимо разработать быстросходящийся алгоритм для нахождения наилучшего соответствия последовательности параметров предъявленной реализации X эталонным последовательностям параметров словаря $\{Y_k\}$, минимизируя величину

$$d = \sum_p \min_k \left(SL^x_p \; \# \; SL_k \right), \qquad (6)$$

где SL^x_p, SL_k содержат сегменты SG^x_{pi}, SG^y_{kj} соответственно; # – операция сопоставления.

Такая постановка задачи распознавания данной информационной технологии приводит к следующей схеме, которая представлена на рис. 4. Рассмотрим основные процедуры модели распознавания.

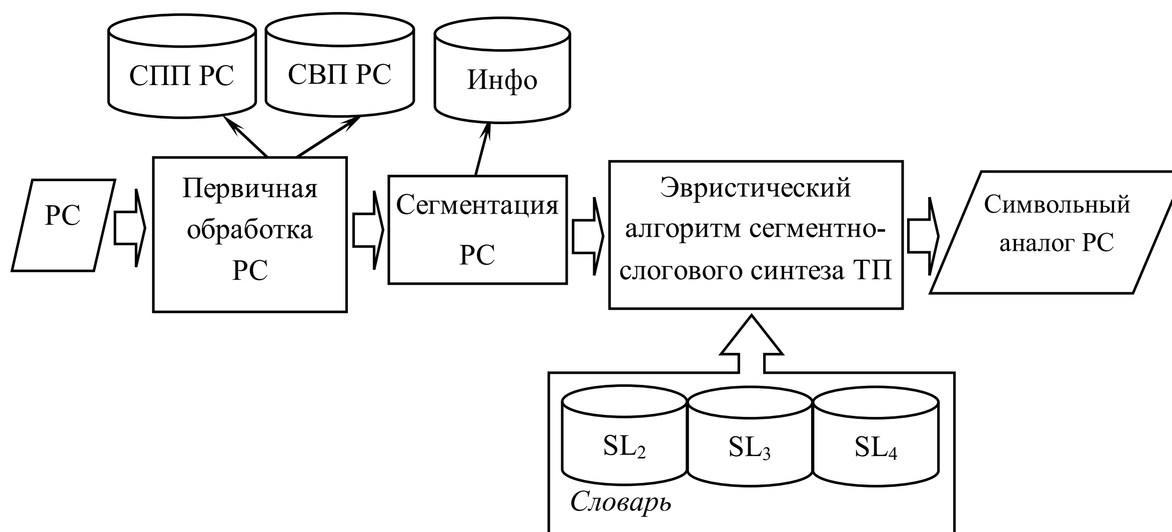


Рисунок 4 – Функционально-структурная схема модели распознавания

Предъявленная реализация РС (с микрофона или из wav-файла) подвергается процедурам первичной обработки и сегментации, которые используются в модели обучения. Сегментированная ТП XE для речевого сигнала, поступающего на вход системы распознавания, рассматривается как совокупность двух-, трех- и четырех-сегментных слогов, для которой выполняется поиск решений-комбинаций для эталонной ТП (ЭТП) с помощью алгоритмов: поиск в ширину и в глубину с использованием эвристических оценок [11], [14], [15]. Для полученного списка комбинаций ЭТП принимается решение о выборе наилучшей по критерию (6). Для синтеза ЭТП выбирается соответствующее описание ТП слогов-эталонов.

Сплайн-синтез ЭТП для СПП. Синтез эталонной ТП XE^* выполняется согласно следующей модели конкатенации ТП слогов-эталонов

$$XE^*(l,t) = \begin{cases} \tilde{Y}_i^1, & N'_0 \le i \le N'_1, \\ \dots \\ \tilde{Y}_i^k, & N'_{k-1}+1 \le i \le N'_k, \\ \dots \\ \tilde{Y}_i^M, & N'_{R-1}+1 \le i \le N'_R, \end{cases} \tag{7}$$

где $k = 1 \div R$; R – количество слогов-эталонов $\tilde{Y}_i^k$ в ЭТП XE; $\tilde{Y}_i^k$ – ТП соответствующей РЕ словаря, восстановленная согласно модели (1), (2); N_k – количество временных отсчетов ТП k-й ТП, $t_1 \in [1, N_1]$, $t_2 \in [1, N_2]$, …, $t_k \in [1, N_k]$, …, $t_R \in [1, N_R]$. Границы слогов внутри текущей комбинации траекторий параметров для ЭТП определяются таким образом:

$N'_0 = 1$; $N'_1 = N_1$; $N'_2 = N_1 + N_2$; $N'_3 = N_1 + N_2 + N_3$; …; $N'_R = N_1 + N_2 + \ldots + N_R$.

Параметры моделей описания траекторий параметров РЕ, составляющих текущую комбинацию ЭТП, уточняются при условии минимизации ошибки в каждой частотной полосе l:

$$y_l^{\,2} \times \sum_{k=1}^{R} \left(\sum_{i=N'_{k-1}}^{N'_k} \left| \tilde{Y}^i_{k,l} - Y^i_{k,l} \right|^2 \right) \to \min . \tag{8}$$

Синтез ЭТП в классе колоколообразных функций для СВП. ТП j-й РЕ внутри текущей комбинации ЭТП определена в прямоугольной области D^j: $[\omega_0,\ \omega_M] \times [t_0,\ t_{Nj}]$ и восстановлена в виде суперпозиции произведения колоколообразных функций $Zt_{(i)}(t_l)$, $Z\omega_{(i)}(\omega_k)$ ($i = 1...L_j$, $k = 1...M$, $l = 1...Nj$). Для текущей комбинации ЭТП $XA^*(\omega, t)$, состоящей из R РЕ словаря, область определения с учетом объединения частотно-временных диапазонов каждой РЕ в комбинации – D: $[\omega_0,\ \omega_M] \times [t_0,\ t_N]$, где $t_N = t_{N_1} + t_{N_2} + \ldots + t_{N_R}$.

Описание СВП для эталонной ТП $XA^*(\omega, t)$ в точке (ω_k, t_l) области определения D вычисляется как суперпозиция R гладких функций $YA^*_m(\omega, t)$ ($m = 1...R$), которые являются аналитическим описанием СВП речевых единиц словаря, таким образом

$$XA^*(\omega_k,\ t_l) = \sum_{m=1}^{R} YA^*_m(\omega_k,\ t_l) = \sum_{m=1}^{R} \left(\sum_{i=1}^{Lm} Z\omega_{(i)}(\omega_k) \cdot Zt_{(i)}(t_l) \right), \tag{9}$$

где R – количество слогов в ЭТП $XA^*(\omega, t)$, L_m – количество параметров колоколообразных функций $Zt_{(i)}(t_l)$, $Z\omega_{(i)}(\omega_k)$ ($i = 1..L_m$) для соответствующего m-го слога комбинации для ЭТП $XA^*(\omega, t)$.

Композиция символьного аналога РС. После завершения работы алгоритма поиска ЭТП, которая наилучшим образом соответствует последовательности параметров предъявленного РС, выполняется композиция ее символьного аналога

$$W = f_{compose}(\mathrm{SL}_1,\ \mathrm{SL}_2,\ \ldots,\ \mathrm{SL}_i,\ \ldots,\ \mathrm{SL}_R).$$

3. Программная реализация информационной технологии

Представленная в статье информационная технология была реализована в виде компьютерной системы распознавания речи *SPeach*. Основной средой, в которой реализованы алгоритмы данной информационной технологии, является Borland Delphi 5. Структурно система *SPeach* реализована в модулях, которые перечислены в табл. 1. Интерфейс программы представлен на рис. 5.

Таблица 1 – Перечень основных модулей системы *SPeach*

Название модуля	Назначение
WAVReadWrite	ввод РС с микрофона или wav-файла
SpectrAnalis	спектральный анализ РС, вычисление параметров
Filters	фильтрация и сглаживание данных
Segmentation	сегментация РС, определение типов сегментов
UnitTeachProcFirmirSlovar	формирование словарей РЕ
ApproximationProc	построение сплайн-описания СПП
Lokony_Anjezi	построение описания СВП в классе колоколообразных функций
Recogn	процедуры ДП, добавление элементов в словарь
PoiskVShirinu	алгоритмы BFS
PoiskVGlubinu	алгоритмы DFS
Evristics	эвристический алгоритм
ModDraw3dGraph	графика

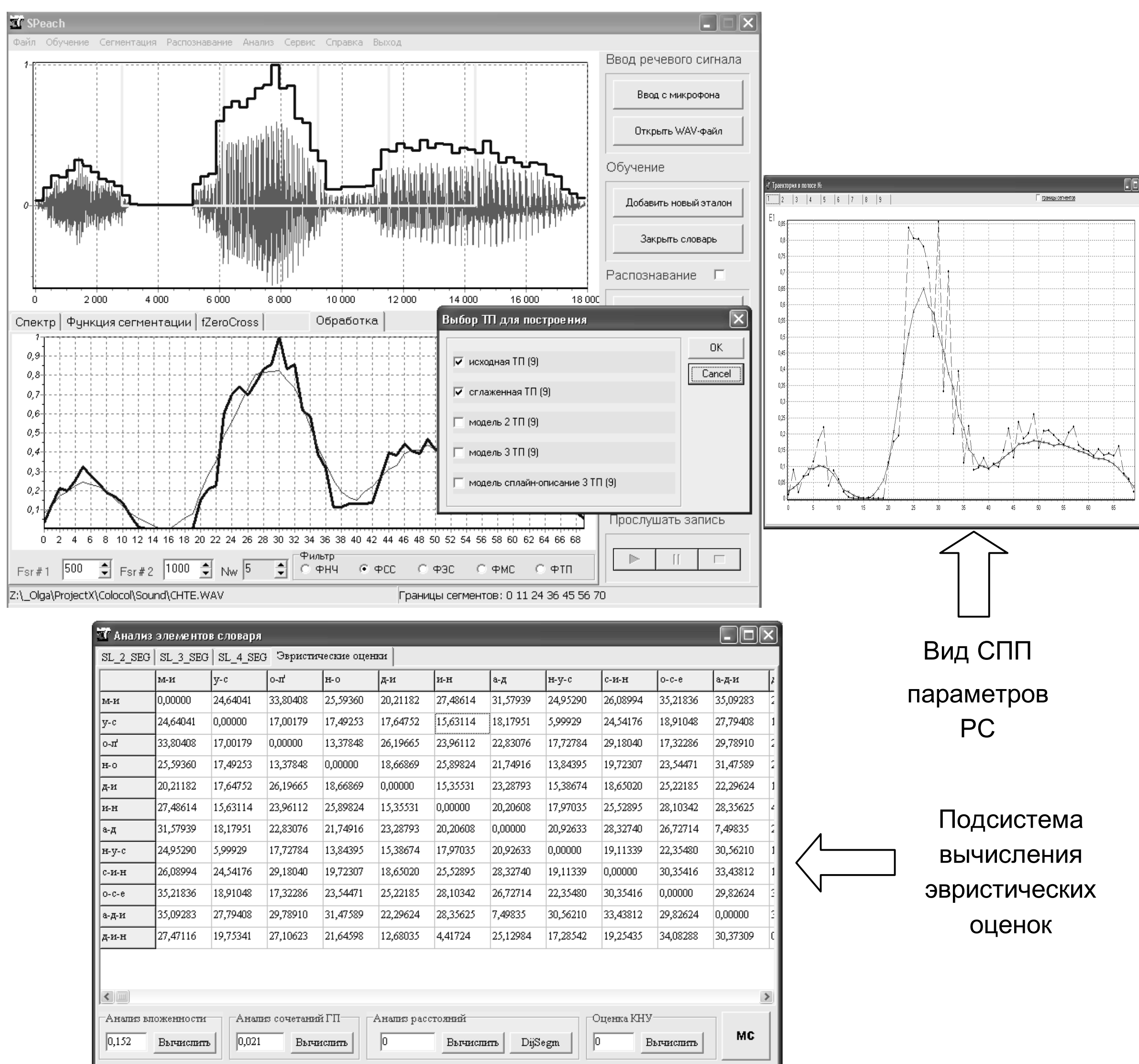


Рисунок 5 – Интерфейс приложения *SPeach*

На рис. 6 а, б, в приведен пример синтеза ЭТП для предъявленной реализации РС «один» согласно представленной модели распознавания.

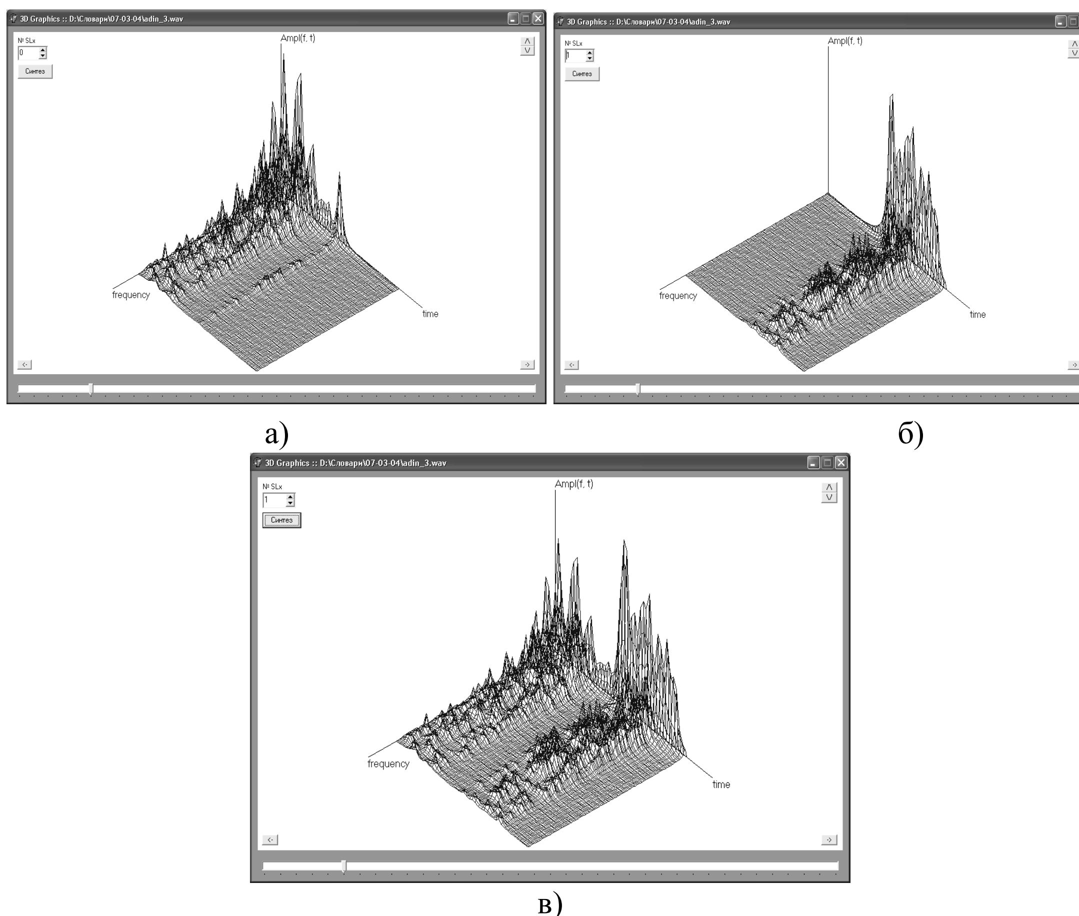


Рисунок 6 – Описание СВП в классе колоколообразных функций:
а) слог «ад»; б) слог «ин»; в) конкатенация слогов «ад» и «ин» в ЭТП

Результаты тестирования разработанной системы *Speech* подтверждают эффективность предложенной информационной технологии.

Выводы

В данной статье предложена новая информационная технология для построения интеллектуальной системы распознавания речи, которая обеспечивает комплексный подход, учитывающий взаимосвязи между иерархией представлений о речевом сигнале. В дальнейшем предполагается усовершенствование эвристического алгоритма поиска с целью повышения быстродействия распознавания.

Литература

1. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – К.: Наукова думка, 1987. – 264 с.
2. Survey of Russian Recognition Systems / A.L. Ronzhin, R.M. Yusupov, I.V. Li, A.B. Leontieva // In Proc. of Int. Conf. SPECOM'2006. – St. Petersburg, 2006. – P. 54-60.

3. Ронжин А.Л., Карпов А.А., Ли И.В. Система автоматического распознавания русской речи SIRIUS // Искусственный интеллект. – 2005. – № 3. – С. 590-601.
4. Obtaining candidate words by polling in a large vocabulary speech recognition system / L.R. Bahl, R. Bakis, P.V. De Souza, R.L. Mercer // In Proc. ICASSP'88 Int. Conf. Acoust., Speech and Signal Process. – NY, 1988. – P. 489-492.
5. The SSI large-vocabulary speaker-inependent continuous speech recognition system / Meisel W.S., Anikst M.T., Pirzadeh S.S., Schumacher J.E., Soares M.C., Trawick D.J. // In Proc. ICASSP'91 Int. Conf. Acoust. – Speech and Signal Process., Toronto, 1991. – P. 337-340.
6. Building acoustic models for a large vocabulary continuous speech recognizer for Russian / M. Tatarnikova, I. Tampel, I. Oparin, Yu. Khokholov // In Proc. of Int. Conf. SPECOM' 2006. – St. Peterburg, 2006. – P. 83-87.
7. Кодзасов С.В., Кривнова О.Ф. Общая фонетика. – М.: РГГУ, 2001. – 592 с.
8. Рассел С., Норвиг П. Искусственный интеллект: современный подход. – М., 2006. – 1408 с.
9. Шаповалова С.И. Оптимизация решения задач поиска на прологе // Искусственный интеллект. – 2000. – № 3. – С. 121-127.
10. Paul D.B. Algorithms for an optimal A*search and linearizing the search in the stack decoder // In Proc. ICASSP'91 Int. Conf. Acoust. – Speech and Signal Process. – Toronto, 1991. – P. 693-696.
11. Карпов О.Н. Технология построения устройств распознавания речи. – Д., 2001. – 184 с.
12. Kopeček I. Speech synthesis based on the composed syllable segments // Proc. of the First Workshop on Text, Speech and Dialogue – TSD’98. – 1998. – P. 259-262.
13. Карпов О.Н., Савенкова О.А. Некоторые эксперименты по повышению надежности распознавания слов заданного словаря // Системные технологии. Региональный межвуз: Сб. науч. тр. – Выпуск 6 (35). – Днепропетровск, 2004. – С. 60-66.
14. Карпов О.Н., Савенкова О.А. Распознавание речи на основе сегментно-слогового синтеза в терминах пространства состояний // Искусственный интеллект. – 2006. – № 3. – С. 532-536.
15. Карпов О.Н., Савенкова О.А. Эвристический алгоритм поиска оптимальных решений сегментно-слогового синтеза // Искусственный интеллект. – 2007. – № 4. – С. 378-385.
16. Осадчий А. Автоматизация проектирования, спецификация ПО. На пути к применению UML // Компьютеры + Программы. – 2002. – № 9 (94). – С. 46-49.
17. Карпов О.Н. Вычислительные схемы представления функций многих переменных в классах функций меньшего числа переменных: Аналитическое описание поверхностей и спектров речевых сигналов: Монография. – Д.: Изд-во Днепропетр. ун-та, 2003. – 120 с.

О.О. Савенкова, О.М. Карпов

Технологія побудови інтелектуальної системи розпізнавання мовлення

У статті розглянуто інформаційну технологію для системи розпізнавання мовлення з великим словником на основі сегментно-складового синтезу траєкторії параметрів. В основу технології покладені моделі навчання та розпізнавання, які використовують швидкозбіжні алгоритми пошуку.

O.A. Savenkova, O.N. Karpov

Information Technology of Design the Intelligent Speech Recognition System

New information technology for the speech recognition system by segment-syllabic parameters trajectory synthesis is considered. The technology is based on convergence algorithms and consists of two models: learning model; recognition model.

Статья поступила в редакцию 17.08.2008.